Detecting Mother-Father Differences in Spending on Children: A New Approach Using Willingness-to-Pay Elicitation

Rebecca Dizon-Ross
University of Chicago

Seema Jayachandran
Princeton University

July 22, 2022

Abstract

This paper tests whether mothers and fathers differ in their spending on their daughters relative to their sons. We compare mothers' and fathers' willingness to pay (WTP) for specific goods for their children, diverging from the previous literature's approach of comparing the expenditure effects of mothers' versus fathers' income. Our method, which we apply in Uganda, allows us to estimate gender differences and explore mechanisms with greater precision. A second innovation is that we examine *why* spending patterns differ between mothers and fathers, e.g., altruism, personal returns to investing in children. We find that fathers have a lower WTP for their daughters' human capital than their sons' human capital, whereas mothers do not. We also find evidence that altruism plays a role in the mother-father differences: fathers' WTP for goods that simply bring joy to their daughters is lower than their WTP for such goods for their sons, but mothers' is not.

*We thank Rebekah Chang, Jake Gosselin, Suanna Oh, and Varun Satish for excellent research assistance and Lori Beaman, Kate Orkin, Heather Sarsons, and Alessandra Voena for helpful comments. Jayachandran gratefully acknowledges support from the National Science Foundation (SES-1156941). Contact information: rdr@chicagobooth.edu; seema@alum.mit.edu.

1 Introduction

Do fathers invest less in their daughters than their sons? Are mothers less discriminatory against their daughters? If so, these relationships would be important for policy, as they would imply that improvements in gender equality are self-reinforcing. As women gain more say in household decision-making, household spending on daughters may increase, producing more gender equality in the next generation. This virtuous cycle could help to close the gender gaps in schooling and health care that are pervasive in developing countries (Evans et al., 2021).

In this paper, we examine if and why fathers underspend on their daughters' health and education relative to how mothers spend. We adopt a new approach to measure parents' spending preferences that has higher statistical power than the traditional approaches used in the literature: we elicit and compare mothers' and fathers' willingness to pay (WTP) for various goods for their sons and daughters. We conduct the study in Uganda among a sample of 1,084 households with young children in which we interviewed one randomly selected parent or usually both (separately).

We find that fathers have a significantly lower WTP for their daughters' human capital than their sons' human capital. In contrast, mothers, if anything, have a higher WTP for their daughters' human capital than their sons'. As a result, willingness to spend on daughters is higher among mothers than fathers. While previous papers have hinted that mothers might spend more on daughters while fathers spend more on sons (see Lundberg, 2005 for a review), the previous literature's power to statistically reject equality between male and female caregivers has been limited (e.g., Duflo (2003)).

Another contribution of our study is to investigate the underlying reason for these mother-father differences. In a review paper on possible mother-father differences in son-daughter preferences published two decades ago, Lundberg (2005) noted that "one central unanswered question for economists is whether these differences emerge from parental preferences ...[or from] real or perceived differences in the returns to parental inputs." Today, we still know remarkably little about this question.

The preference-based explanation is simple: fathers care more about their sons than their daughters, and mothers do not. Alternatively, the differences between mothers and fathers could be due to different returns to investment ("investment-based" explanation). Mothers might expect to be more dependent on support from their children in old age, as

women have lower earnings and longer life expectancy than men, on average. This could cause mothers to spend more on their daughters than fathers do if they believe, as the majority of our sample does, that daughters are more likely to help support their parents in old age than sons. Fathers and mothers could also have different beliefs about the returns to human capital investment for boys and girls.

To test between these hypotheses, we examine whether there are similar mother-father son-daughter WTP differences for goods that bring joy to the children but do not add to their human capital: toys and candy. Under an investment-based explanation, one would expect the gaps to be observed for human capital goods but not toys and candy. Conversely, the patterns being similar for both types of goods would support a preference-based explanation.

The evidence supports a preference-based explanation: We find that fathers also have a lower WTP for goods that bring joy to their girls than to their boys, suggesting that they have less altruism or love for their daughters than their sons. Mothers, in contrast, have no lower WTP for goods that bring joy for their girls than for their boys.

To further probe the different explanations, we collect data on which parent the respondents view as caring about the children more. We find that the mother-father differences are driven entirely by households where both parents believe the mother loves the children more than the father does, which could proxy for loving girls more than the father does in a context where everyone might love their sons. We also collect data on beliefs about the returns to investing in children and examine heterogeneity in the WTP differences based on these beliefs. We do not find evidence for investment-based explanations, although we cannot reject that they play a role as well.

Our paper makes two contributions to the literature on how parents spend on their children. The first is to propose and implement a way of measuring parents' spending preferences that provides enough statistical power to test if mothers and fathers have the same spending patterns for their sons and daughters. The method has high statistical power because it zeros in on the expenditure category of interest.

The standard approach in the literature is to examine the effects of exogenous changes in male versus female income, asking: does a change in income for mothers lead to larger spending on, say, girls' education, or children's education in general, than the same change in income for fathers? However, children's health and education — and goods assignable to children more generally — typically constitute a modest share of household expenditures. Detecting mother-father differences off a small base is statistically challenging. Thus, while

a few studies can reject that women's income and men's income have identical effects (e.g., Lundberg et al., 1996) or can rule out modest gender differences (e.g., Somville et al., 2020), many studies in this literature are underpowered to compare mothers and fathers spending, let alone whether mother-father patterns differ for sons versus daughters. We are unaware of any prior paper that uses this approach and statistically rejects that mothers and fathers spend identically on their sons relative to daughters.[1,2]

In contrast, our method allows us to establish that mothers discriminate significantly less against daughters than fathers do in their human capital spending. This finding provides important, policy-relevant evidence, supporting the idea that female empowerment might be self-reinforcing across generations.

To be more concrete about the statistical power advantages, Haushofer and Shapiro (2016) study large cash grants in Kenya and report that the minimum detectable effect (MDE) size gap between male and female grant recipients for their health and education outcomes is 0.24 to 0.25 standard deviations (SD). The MDE for mother-father differences in child spending in our study, which uses a comparable number of households, is 0.08 SD.

Another advantage of using WTP to study parental differences in spending is that it directly elicits parent-specific choices rather than inferring them from household-level choices. In the standard approach, if an increase in women's relative income does not change household spending on children, that might be because women's bargaining power did not increase, in which case one cannot conclude whether or not mothers and fathers have similar spending patterns.

Finally, the approach we use is practical. It could be easily incorporated into surveys that are interested in measuring parental spending or son preference. The ability to compare WTP for goods with different attributes or goods relevant for different children offers valuable flexibility to test mechanisms.

The second way we advance the literature is by investigating *why* mothers and fathers

[1]For example, Duflo (2003) analyzes the child health effects of pensions in South Africa, and one likely could not reject that there is no difference in the effect of grandmothers' versus grandfathers' income on boys relative to girls. (The paper does not report this test.) Other studies on male-female differences in investment in children include Thomas (1994), Qian (2008), Benhassine et al. (2015), and Akresh et al. (2016).

[2]Two related papers use lab experiments to compare mothers' and fathers' allocation of money to their child or to themselves (Cherchye et al., 2021; Ringdal and Sjursen, 2021). These studies were conducted after ours (and cite our study). Ringdal and Sjursen (2021) also find some evidence for gender-concordant patterns. Another related study is Nikiforidis et al. (2018), who asked a small sample of mothers and fathers visiting a zoo in the US to choose between a boy's or girl's backpack and posed a similar question about a savings bond to an Amazon MTurk sample, with the main finding that choices were gender-concordant.

spend differently on their children. One related paper is Doepke and Tertilt (2019), which presents and tests a model in which mothers and fathers have identical preferences, but mothers spend more on children's human capital because each parent specializes in providing different public goods for the household. We add to this quite small literature.

2 Data

The data for the study were collected in Iganga district in eastern Uganda. The sample comprises households with a child in primary school. In the first round of data collection (March-May 2013) we surveyed one randomly-selected parent per household. The randomization means that household and child characteristics are balanced when we compare mothers and fathers. In the second round of data collection (September-October 2013), we returned to a subset of the households, specifically those who also had a child age 3 to 8 years old and surveyed the other parent.

Sampling We sampled households with children enrolled in grade 4 to 6 in 40 government primary schools with whom we partnered to offer one of our education goods (practice exams that schools administer for a fee). We began with a listing of eligible households in the participating schools. The first eligibility criterion was that the child lived with both biological parents (94% of children). The second was that, for the current academic term, the child's parents had not paid for all of the practice exams the school offered (70% of children). The rationale was that we would be eliciting WTP at different prices up to but not exceeding the market price, so households already purchasing the good at the market price would generate no variation in WTP.

First survey For the first survey, we randomly selected whether to interview the mother or the father, stratified on school and whether the household was polygamous. A surveyor visited the home and administered a screening questionnaire to confirm eligibility. The final sample comprises 1084 households that met the eligibility criteria and agreed to participate.

The survey elicited WTP for goods for the focal child enrolled in grade 4-6.[3] If a household also had a younger child aged 3-8, we gathered WTP for one good for that younger child. Hereafter, we refer to these two children as the older and younger child, respectively.

The randomly chosen parent was interviewed with no one else present, besides infants or toddlers. The survey first collected information on household composition, family back-

[3]In cases of more than one eligible older child in a household, we randomly selected one, and the same for the younger child.

ground, and income and assets for each parent. Second, the survey elicited WTP for a set of goods. Finally, the survey asked questions related to mechanisms, such as expected old age support. The survey took approximately 75 minutes to complete. The participant received 8000 or 10,000 Ugandan shillings (1 USD $\approx$ 2600 UGX) as compensation for their time and to minimize cash constraints affecting WTP.[4]

Second survey In the second round of data collection, we revisited 729 of the original 1084 households and surveyed the other parent, using a similar survey structure and content. The reason for revisiting a subsample was budgetary constraints. We focused on the subsample with a child aged 3 to 8. Because one of the goods we used was shoes for the young child, we over-sampled households in which the child did not own shoes, according to the parent interviewed in the first survey. We attempted to interview all 704 of these households and completed interviews with 646. We also resurveyed a random subset (83 households) of the other households with a child age 3 to 8. All respondents received 9000 UGX for participating.

Procedure for WTP elicitation

To elicit WTP, we used the Becker-DeGroot-Marschak (BDM) mechanism, asking the respondent if he or she was willing to purchase the good at a series of prices, in declining order from the market price to a price near zero (Becker et al., 1964).[5] The decrement was chosen so that respondents were asked about roughly 12 price levels per good. The respondent was told that after the price questions, one price would be randomly chosen and she would purchase the good from us at that price if and only if her response had been that she wanted to. The surveyor explained the procedure in detail to ensure comprehension, and we also asked debrief questions (such as regret about one's choices) to confirm comprehension. The selection of the randomized price and exchange of money and goods, if applicable, were conducted just after the BDM questions were asked for a good. In the second wave, to increase sample size without increasing study costs, we grouped five of the goods, and first randomly chose one good and then one price level for that good; respondents were informed in advance that a transaction could only occur for one of these five goods; this two-step randomization occurred after the BDM questions for all five goods.

[4]We randomized the payment level to test for cash-on-hand and gift-exchange effects on WTP. Appendix Table A.1 shows that receiving the higher compensation level does not affect WTP for the goods we offered.

[5]Recent studies validating the use of BDM in developing countries include Berry et al. (2020) and Burchardi et al. (2021).

Also to increase sample size, in each wave we asked WTP in a similar but non-incentivized way for additional goods. The surveyor followed the same protocol of showing the actual good to the respondent so that it was concrete, but respondents knew in advance that for these goods, no transaction would take place. Later in this section, we present evidence that the non-incentivized WTP elicitation appears to have worked very similarly to the incentivized WTP elicitation. As a result, we pool incentivized and non-incentivized WTP in our main specifications for statistical power.

Children's goods

We used several criteria when choosing which goods to offer parents. First, we wanted parents to be familiar with the good and its market price; otherwise, based on piloting, variation in the perceived quality and market price would add noise and potentially bias the results. Second, the good should be something that most households value at less than the market price, but place some value on, so that there is variation in WTP. Third, the good should not be particularly appealing to one gender, *within* the categories of human capital or "enjoyment goods" (non-human capital); the goods are intended to represent the broad categories of human capital or enjoyment goods, so while each whole category might be favored more by one gender, we would not want a good that, say, fathers idiosyncratically like more than mothers do or that is more appropriate for girls than boys. In addition to doing extensive preliminary fieldwork to choose goods that met these criteria, we asked questions on the survey to verify our assumptions.

Appendix Table A.2 lists each child good, which survey it was collected in, which focal child it was for, whether the WTP elicitation was incentivized, and the market price. Prices ranged from 1500 to 6000 UGX.

Human capital goods In the first survey, we elicited WTP for three human capital goods. The first, measured in an incentivized way, was practice exams administered by the child's school. Schools offer practice exams during and at the end of each of the three terms of the school year, but charge students to participate. Our survey was conducted during Term 1 of 2013, and we sold a bundle of all of the exams for Terms 2 and 3.

The other human capital goods in the first survey were deworming medicine for the older child and, if the household had a child age 3 to 8, rubber-soled shoes for them.[6] Many

[6]In the first survey, we asked about deworming only for those randomized to receive the higher payment for participating in the study. The reason was to justify the higher compensation for some people by their survey being longer.

young children do not wear shoes, and being barefoot is a risk factor for intestinal worms, as well as cuts and injuries. WTP for these two goods was measured in a non-incentivized way.

In the second survey, we elicited WTP for four human capital goods: rubber-soled shoes for the young child (incentivized); a grade-appropriate math workbook for the older child (incentivized); deworming medicine for the younger child (non-incentivized) and practice exams for the older child for the first two terms of the 2014 school year (non-incentivized).

Enjoyment goods In the second survey, we also elicited WTP for fun goods for children that are not human-capital enhancing. Both goods were offered for the younger child: a rubber ball (incentivized) and a packet of candy (incentivized).[7]

Benchmark goods

We elicited WTP for goods used by adults to "benchmark" each respondent's general WTP for goods. We control for the adult good WTP in our regressions to increase precision, as factors such as cash on hand or gift-exchange motives should affect adult good WTP similarly to how they affect child good WTP (Dizon-Ross and Jayachandran, forthcoming, 2022). In principle, controlling for the adult good WTP could also help address any systematic differences in such factors between male and female respondents. In practice, controlling for the adult good WTP does not change our point estimates much, only the standard errors, and we show robustness to excluding the control. Through piloting, we identified adult goods that were not gendered. In the first survey, the adult good was a metal cup for drinking (3600 UGX) and in the second survey, the two adult goods were a poster (2000 UGX) and a pair of jerry cans (4000 UGX).

Validation of non-incentivized WTP elicitation

Appendix B presents a validation of the use of non-incentivized WTP. The validation analysis uses practice tests, a good for which we measured WTP in both an incentivized (first survey) and non-incentivized (second survey) manner.

We use LASSO to separately identify which variables predict the incentivized WTP and which predict the non-incentivized WTP, from among a large set of household characteristics and survey responses (e.g., perceived quality of tests) that we asked in both survey rounds.

[7]We also asked about a separate toy for girls (teddy bear) and boys (toy truck) in both surveys (non-incentivized). To make comparisons across boys and girls, it is important to use the same gender-neutral enjoyment good, so we exclude these toys from our analysis.

We find that LASSO chooses the same predictors for incentivized and non-incentivized WTP. In addition, the predictive coefficients are quantitatively similar, and we cannot reject identical predictive relationships.[8]

Survey questions to test mechanisms

We also asked explicit survey questions about the potential mechanisms we focus on. Our strongest test of the hypothesis that mothers and fathers differ in how much they care about their children's well-being is to examine WTP for enjoyment goods, but we also asked direct questions about which parent cared more about the children. A second explanation we consider is that mothers and fathers differ in how much they will personally benefit from having higher human capital children, specifically because of financial support from them. We asked whether the respondent thought they or their spouse would be more dependent on financial support from children and how expected support differed by child gender.

Summary statistics and balance

For 729 households, we have surveys of both the mother and father. For 355 additional households that we did not revisit in the second round, we have data for one randomly-chosen parent. To verify that the randomization yielded balance, Table A.3 compares summary statistics for household and focal child characteristics in the mother and father samples. Almost all households in our sample own land, and 25% are polygamous. Older focal children are 12 years old, on average, and younger focal children, almost 6.

We conduct an omnibus balance test between the mother and father subsamples. We fail to reject the null of joint orthogonality of all variables (p-value 0.35). Following Imbens and Rubin (2015), we also calculate the difference between groups divided by the pooled standard deviation. These standardized differences are all far below the rule-of-thumb "cutoff" for good balance of 0.25 SD. Table A.4 shows that, in addition, within the subsamples of male focal children and of female focal children, mothers and fathers have balanced characteristics.

While we randomized the gender of the parent within each household, we did not randomize child gender. Reassuringly, Table A.5 shows that household and parent characteristics are nevertheless similar between the girl and boy subsamples. An omnibus balance test fails to reject the null that they are identical, and all standardized differences between the

[8] As another way to assess the non-incentivized elicitation, in the first survey we elicited non-incentivized WTP for a good chosen to be gendered, a kitchen sieve. Reassuringly, women have a 0.14 standard deviation higher WTP for the sieve than men do. (We do not use WTP for the sieve in our main analyses.)

two samples are far below 0.25 SD. Appendix Table A.6 shows that there is also boy-girl balance within the subsamples of mothers and of fathers.

A parent's gender is bundled with other individual characteristics, such as earnings, and child gender is similarly bundled with other traits. Tables A.7 and A.8 summarize the personal characteristics of mothers and fathers, and of daughters and sons, respectively. Mothers are younger and have less income than fathers, on average. In contrast, daughters and sons have similar characteristics, such as age and school performance.[9]

3 Empirical strategy and results

Spending on children's goods

We begin by testing whether parents collectively spend more on their daughters or sons and whether mothers or fathers spend more on average on their children. To do so, we estimate the following equation:

$$WTP_{ihgc} = \alpha + \beta Daughter_{ihc} + \gamma Mother_{ihc} + \delta X_{ihgc} + \epsilon_{ihgc} \quad (1)$$

where each observation is for parent i in household h asked about a good g for child c. The independent variables of interest are $Daughter$, an indicator for the child being female and $Mother$, an indicator for the respondent being female. In principle, we could estimate the difference using a single good, but for statistical power and so that the results are less specific to a particular good, we pool several goods. To make WTP comparable across goods, we normalize the WTP for each good by its within-sample standard deviation, with the results robust to other normalizations. The vector of control variables X includes good-by-survey-wave fixed effects, and stratum fixed effects. To increase precision, X also includes WTP for the adult goods (for which men and and women have similar preferences). The standard errors allow for clustering within a household.

Column (1) of Table 1 presents the results. There is no statistically significant difference between parents' WTP for goods for their daughters versus their sons and also no statistical difference between mothers' and fathers' WTP overall.

Next, to understand whether daughter-son spending preferences differ across mothers

[9]We have relatively more older focal children among girls (standardized difference of 0.185), likely due to sampling error. We show that all of our results are robust to controlling for whether the child is the younger focal child in parallel to how child gender enters the regressions.

Table 1: Fathers spend less on girls but mothers do not

	(1)	(2)	(3)	(4)
Daughter	-0.035	-0.102***	-0.069*	-0.095**
	[0.024]	[0.032]	[0.036]	[0.037]
Mother × Daughter		0.135***	0.079	0.147***
		[0.047]	[0.053]	[0.053]
Mother	-0.041	-0.109***	-0.111***	-0.076*
	[0.028]	[0.036]	[0.040]	[0.043]
p-val: Mother + Mother × Daughter = 0		0.461	0.444	0.092
p-val: Daughter + Mother × Daughter = 0		0.328	0.813	0.180
Dep. var. mean father-son	1.979	1.979	1.841	1.979
Fixed effects	Stratum	Stratum	Stratum	HH
Goods included	All	All	Incentivized	All
Number of observations	6,687	6,687	4,000	6,687

Notes. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All columns control for survey round, adult WTP, adult WTP interacted with survey round, and good fixed effects. Columns 1-3 control for strata fixed effects; Column 4 controls for household fixed effects. Standard errors are clustered by household.

and fathers, we add the interaction term and estimate the following equation:

$$WTP_{ihgc} = \alpha + \beta Daughter_{ihc} + \delta Mother_{ih} \times Daughter_{ihc} + \lambda Mother_{ih} + \gamma X_{ihgc} + \epsilon_{ihgc} \quad (2)$$

In this case, the coefficient on *Daughter* tells us how much lower fathers' WTP is for daughters than sons, and the coefficient on $Mother \times Daughter$ tells us how different the *Daughter* effect (i.e., the spending on daughters relative to sons) is for mothers than fathers. Finally, the coefficient on *Mother* shows us how much less mothers spend overall (on sons) than fathers.

Column (2) of Table 1 shows that fathers have a lower WTP for goods for their daughters while mothers do not. The coefficient on *Daughter* is negative and significant, suggesting that fathers have a 0.10 SD lower WTP for daughters than sons. The coefficient on $Mother \times Daughter$ is positive, significant, and similar in magnitude to the *Daughter* coefficient, which means that the net effect of *Daughter* for mothers is 0. At the bottom of the table we report the p-value for whether mothers spend equally on daughters and sons ($Daughter + Mother \times Daughter$).

Note that mothers' spending level for sons is lower than fathers', as seen from the main effect of *Mother*. Combining this fact with the positive *Mother* × *Daughter* effect, on net, mothers and fathers spend an indistinguishable amount on daughters. (The bottom of the table reports the p-value for this test, *Mother* + *Mother* × *Daughter*). Thus the key difference between fathers and mothers is that only fathers spend more on their sons than their daughters. The lower average spending of mothers is consistent with women having less control over household income than their husbands.

The findings reported in column (2) are robust across several different specifications. Column (3) verifies the robustness to limiting to the goods where we incentivized the WTP elicitation. The precision of the estimates falls, and the magnitude of the coefficients decreases somewhat, but the qualitative takeaway is the same: fathers have significantly lower spending on daughters than sons, whereas mothers do not. Column (4) shows that the results are robust to including household fixed effects.

Appendix Table A.9 shows that the findings are robust to normalizing the WTP for each good by its sample mean rather than sample standard deviation. The main effect of -0.06 for *Daughter* indicates that fathers have 6% lower WTP for goods for their daughters than for their sons. Since daughters are marginally more likely than sons to be the older focal child, Appendix Table A.10 shows that the results are robust to controlling for an indicator that the child is the younger focal child in parallel with controlling for *Daughter*. Finally, Appendix Table A.11 shows that our results are robust to excluding the control for the respondent's WTP for the adult good; the main change is that the standard errors are about 30% larger when we omit this control variable.[10]

Spending on human capital

The child development implications of parents' spending preferences depend on whether the results from Table 1 hold for human capital in particular. Column (1) of Table 2 estimates equation (2) using human capital goods only. Mean WTP for each human capital good is 2,000 UGX, roughly 15% of total per-term per-child educational spending.

Again, fathers spend significantly more on sons than daughters, with an effect size of -0.11 SD. Meanwhile, mothers, if anything, spend more on daughters than sons, as *Daughter* + *Mother* × *Daughter* is positive. As a result, mothers spend .07 SD more on girls than fathers do (p-value for *Mother* + *Mother* × *Daughter* is 0.056). Figure 1(a) depicts the

[10] Appendix Tables A.12 to A.14 repeat these robustness checks for the results reported in Table 2.

Table 2: Fathers spend less than mothers on daughters' human capital

	Goods included		
	Human capital	Education	Health
	(1)	(2)	(3)
Daughter	-0.105***	-0.081	-0.138***
	[0.036]	[0.051]	[0.052]
Mother × Daughter	0.148***	0.085	0.200***
	[0.051]	[0.070]	[0.075]
Mother	-0.074*	-0.064	-0.073
	[0.039]	[0.052]	[0.052]
p-val: Mother + Mother × Daughter = 0	0.055	0.639	0.023
p-val: Daughter + Mother × Daughter = 0	0.254	0.930	0.243
Dep. var. mean father-son	2.042	1.618	2.380
Number of observations	5,229	2,542	2,687

Notes. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All columns control for strata and good fixed effects, survey round, adult WTP, and adult WTP interacted with survey round.

estimates graphically.

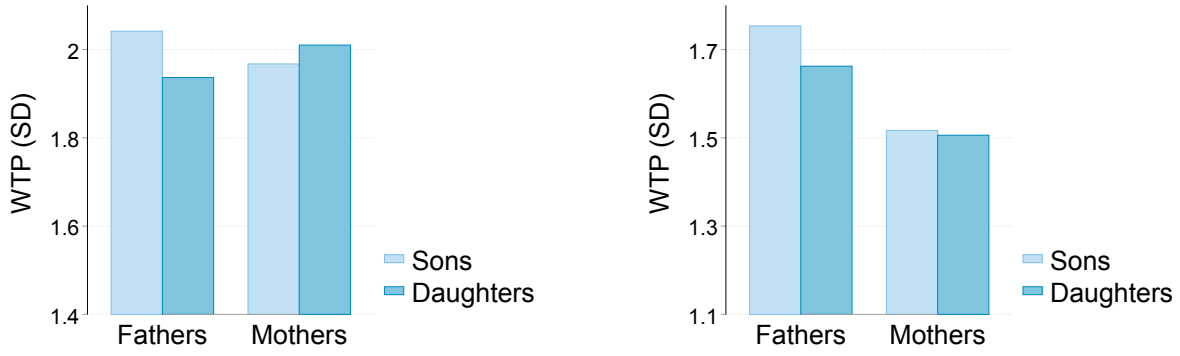
Columns (2) and (3) of Table 2 show that the patterns are qualitatively similar for the two subcategories of human capital – education and health – with the magnitude and significance of the coefficients larger for health.

Testing why mothers and fathers spend differently

What underlies these differences between mothers and fathers in their spending on sons versus daughters? We begin by testing whether altruism towards daughters relative to sons plays a role. We examine parents' WTP for goods that are purely for their children's enjoyment – and do not build human capital. Differential WTP for those goods by child gender suggests differences in altruism towards children. Appendix Figures A.1 and A.2 verify that the benefit that parents thought that these goods had for their children was mostly joy.[11] While these goods could also distract the child, this is a negligible reason stated for the rubber ball, and has no significant gender differences for candy.

[11]Fathers are moderately more likely to see the rubber ball's primary value as happiness for their girls than their boys, but Appendix Table A.15 shows that the results presented next are also robust to controlling for fixed effects for the parent's main reason for valuing each good, so the results are not driven by differences across parent gender or child gender in how parents perceive the candy and rubber ball.

Figure 1: For both human capital and enjoyment goods, fathers spend less on daughters while mothers do not



(a) Human capital

(b) Enjoyment goods

Notes. Figures 1(a) and 1(b) present the coefficient estimates from estimating equation 2 using only human capital goods and only enjoyment goods, respectively.

We first estimate equation 2 using only the "pure enjoyment" (non-human capital) goods and display the results in Figure 1(b). Among fathers, WTP for fun items for daughters is considerably lower than for boys, while, among mothers, the gap is negligible.

To formally test for the differences in patterns between human capital and non-human capital, we estimate equation 2 with all goods, both human capital and non, and with all regressors interacted with a dummy for the good not being a human capital good. We display the results in Table 3, column (1).

The results show that fathers have a lower WTP for enjoyment goods for their daughters, just as they do for human capital goods. The p-value for this test ($Daughter + Daughter \times NonHumanCapital$) is 0.07. This suggests that they have lower altruism for their daughters than sons. The estimate is statistically indistinguishable from the estimate with human capital goods, as evidenced by the fact that the coefficient for $Daughter \times NonHumanCapital$ in Table 3, column (1) is near 0 in magnitude and statistically insignificant. In contrast, mothers have no lower WTP for their daughter's enjoyment than their sons, as seen in Figure 1(b).[12] Unlike for human capital, the point estimate for mothers' WTP for non-human capital is not actually *higher* for their daughters than sons. However, we cannot reject that it is

[12]Mothers also have markedly lower spending on non-human capital goods for both sons and daughters, as evidenced by the negative $Mother \times NonHumanCapital$ coefficient. While this could reflect mothers having lower altruism overall towards their children, it could also reflect other factors such as mothers not believing in spending on "frivolous" goods. These other factors, however, should not differ between sons and daughters and so should not affect our coefficient of interest, $Mother \times Daughter$.

Table 3: Altruism appears to explain some of mother-father differences

	Covariate is:		
	Enjoyment good	Believe mothers love more	Believe daughters support more and mothers receive more
	(1)	(2)	(3)
Mother × Daughter × Covariate	-0.067	0.228**	0.062
	[0.086]	[0.112]	[0.111]
Mother × Daughter	0.148***	0.029	0.122*
	[0.052]	[0.085]	[0.069]
Daughter × Covariate	0.014	-0.124	-0.106
	[0.057]	[0.079]	[0.084]
Daughter	-0.105***	-0.068	-0.089*
	[0.037]	[0.059]	[0.047]
Mother × Covariate	-0.163**	-0.017	-0.098
	[0.065]	[0.085]	[0.087]
Mother	-0.074*	-0.059	-0.041
	[0.040]	[0.067]	[0.052]
p-val: Mother × Daughter + Mother × Daughter × Cov. = 0	0.070	0.000	0.005
p-val: Daughter + Daughter × Cov = 0	0.312	0.000	0.036
Dep. var. mean father-son	1.979	2.074	2.074
Goods included	All	Hum Cap	Hum Cap
Number of observations	6,687	4,644	4,644

Notes. Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All columns control for strata and good fixed effects, survey round, adult WTP, adult WTP interacted with survey round, and all previous controls interacted with the covariate of interest. Columns 2 and 3 are limited to the households with observations from both the mother and the father. Dependent variable is WTP.

higher. Indeed, Table 3 column (1) shows that we cannot reject that the $Mom \times Daughter$ effect is the same for human capital and non-human capital (p-value = 0.41), and that differential altruism toward girls between mothers and fathers fully underlies mothers' and fathers' differences in their willingness to spend on their daughters.

We next provide a second test of the hypothesis that altruism underlies the mother-father differences in discrimination against daughters. Here we conduct heterogeneity analy-

sis using a proxy for whether the mother is the more altruistic parent within the household. We estimate equation (2) using human capital goods only, and fully interact all regressors with a dummy for whether a household is in the roughly 50% of households where both parents identify the mother as the parent who cares more about the children. (In the other 50% of households, either only one or neither of the parents identified the mother as the more loving parent.)[13] While we did not ask about which parent cared more about their girls in particular, in a context where boys might be more universally beloved, we believe caring more about children in general may also proxy for caring more about female children.

Column (2) of Table 3 shows that mothers' greater WTP for daughters' human capital, compared to fathers, comes fully from the households that believe the mother loves the children more. The $Mother \times Daughter$ coefficient — which is the $Mother \times Daughter$ effect among households who do not believe the mothers love the children more — is small in magnitude and indistinguishable from zero. In contrast, the $Mother \times Daughter$ effect among the households that believe the mothers love the children more is a significant 0.26 SD. In addition, the negative $Daughter$ effect among fathers is three times the magnitude and only significant in the subsample that believes mothers are the more loving parent (p-value for $Daughter + Daughter \times Covariate < .01$).

This evidence suggests that altruism is important in explaining the mother-father differences in son preference. However, altruism is not mutually exclusive with investment motives; do those contribute as well? If so, then we would expect the negative $Daughter$ and positive $Mother \times Daughter$ effects to be concentrated among households where the parents believe that (a) daughters support their parents more than sons, and (b) mothers receive more old-age support than fathers from their children. More parents believe each of (a) and (b) than believe their opposites. We use survey measures to construct an indicator that a parent believes both (a) and (b) are true, which we consider a measure of mothers receiving a higher return on investment (ROI) from investing in daughters than fathers do. We average this measure at the household level and estimate equation (2) fully interacted with a dummy that the household is in the top half of households for this measure ("high-ROI-for-mothers" households). We present the evidence in column (3) of Table 3. While the point estimate are consistent with the ROI hypothesis, with the $Mother \times Daughter$ effect being 0.04 SD more positive in the high-ROI-for-mothers subsample and the $Daughter$ effect

[13] We can only construct this variable for households where we surveyed both parents, and so the regression only includes those households.

0.06 SD more negative, we cannot reject equality of the coefficients across subsamples. We view this as weak evidence that ROI motives might also play a role.

To the extent that investment motives are the "residual explanation" for mother-father differences varying between human capital and non-human capital, the estimates in column (1) are also consistent with investment motives playing a role. The $Mother \times Daughter \times NonHumanCapital$ coefficient is negative, albeit insignificant. The sign of the coefficient suggests that mothers' relatively higher spending on daughters is especially pronounced for human capital, which is what one would expect if there were investment motives. This difference is also visible if one compares Figures 1(a) and 1(b). Mothers spend the same amount on their sons' and daughters' non-human capital, but spend more on their daughters' human capital than their sons'.

To summarize, we find that differential altruism towards daughters is an important contributor to the mother-daughter differences in WTP for human capital of daughters relative to sons, while our evidence is inconclusive on whether investment motives also play a role.

Alternative explanations A variant of the investment story is that mothers perceive the returns to female human capital (relative to male human capital) to be higher than men do. However, as shown in Appendix Table A.16, for several questions about parents' beliefs about the value of schooling for boys and girls, there are no meaningful differences between mothers and fathers.

Another possibility is a norm that mothers spend on girls and fathers spend on boys. However, the vast majority of both mothers and fathers, regardless of the gender of their children, report that the norm is for fathers to pay for human capital. A spending norm would also likely be linked with a decision-making norm, but Appendix Table A.17 shows there is no gender concordance in decision-making about health care for a sick child, for example.

4 Conclusion

We revisit the classic question in family economics of whether mothers and fathers spend differently on children, using a different approach than past studies: we elicit each parent's WTP for spending on goods for their children. The advantages of this approach, compared to using exogenous changes in women's and men's income, are statistical power and the ability to choose goods with attributes that enable one to test mechanisms. We apply this method in rural Uganda.

We find that fathers but not mothers spend less on daughters than sons and then investigate why that is. Specifically, we test between a preference-based explanation, in which mothers care about daughters relatively more than fathers do, and an investment-based explanation, in which mothers enjoy a higher financial return on investment in daughters. We find support for the preference-based explanation. A key test examines parents' WTP for goods that bring joy to the children but do not build their human capital. We find similar patterns for these non-human-capital goods as we did for human capital, consistent with fathers' lower altruism toward their daughters playing an important role in spending differences.

Our investigation leaves many questions unanswered regarding what underlies parental spending differences. We do not consider all possible explanations, and we do not explore the deep causes of preference differences. For example, the literature in sociology and psychology has proposed that preference differences could stem from parents identifying more closely with same-gender children (Belsky, 1979; Nikiforidis et al., 2018). Our results are consistent with men and women both having same-gender favoritism. If mothers and fathers had equal financial resources, such favoritism would cancel out. However, because men control more resources than women do, daughters end up disadvantaged. Continuing to explore the reasons for parental differences in spending is a rich area for further research, and using WTP elicitation as a research design could aid in this exploration.

References

Akresh, R., D. De Walque, and H. Kazianga (2016): "Evidence from a Randomized Evaluation of the Household Welfare Impacts of Conditional and Unconditional Cash Transfers Given to Mothers or Fathers," *World Bank Policy Research Working Paper*.

Becker, G. M., M. H. DeGroot, and J. Marschak (1964): "Measuring Utility by a Single-Response Sequential Method," *Behavioral Science*, 9, 226–232.

Belsky, J. (1979): "Mother–Father–Infant Interaction: A Naturalistic Observational Study," *Developmental Psychology*, 15, 601.

Benhassine, N., F. Devoto, E. Duflo, P. Dupas, and V. Pouliquen (2015): "Turning a Shove Into a Nudge? A "Labeled Cash Transfer" for Education," *American Economic Journal: Economic Policy*, 7, 86–125.

Berry, J., G. Fischer, and R. Guiteras (2020): "Eliciting and Utilizing Willingness to Pay: Evidence From Field Trials in Northern Ghana," *Journal of Political Economy*, 128, 1436–1473.

Burchardi, K., J. De Quidt, S. Gulesci, B. Lerva, and S. Tripodi (2021): "Testing Willingness to Pay Elicitation Mechanisms in the Field: Evidence from Uganda," *Journal of Development Economics*, 102701.

Cherchye, L., P.-A. Chiappori, B. De Rock, C. Ringdal, and F. Vermeulen (2021): "Feed the Children," .

Dizon-Ross, R. and S. Jayachandran (forthcoming, 2022): "Improving Willingness-to-Pay Elicitation by Including a Benchmark Good," *American Economic Association Papers and Proceedings*, 122.

Doepke, M. and M. Tertilt (2019): "Does Female Empowerment Promote Economic Development?" *Journal of Economic Growth*, 24, 309–343.

Duflo, E. (2003): "Grandmothers and Granddaughters: Old Age Pension and Intra-Household Allocation in South Africa," *World Bank Economic Review*, 17.

Evans, D. K., M. Akmal, and P. Jakiela (2021): "Gender gaps in education: The long view," *IZA Journal of Development and Migration*, 12.

Haushofer, J. and J. Shapiro (2016): "The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya," *Quarterly Journal of Economics*, 131, 1973–2042.

Imbens, G. W. and D. B. Rubin (2015): *Causal Inference in Statistics, Social, and Biomedical Sciences*, Cambridge, UK: Cambridge University Press.

Lundberg, S. (2005): "Sons, Daughters, and Parental Behaviour," *Oxford Review of Economic Policy*, 21, 340–356.

Lundberg, S. J., R. A. Pollak, and T. J. Wales (1996): "Do Husbands and Wives Pool Their Resources? Evidence from the United Kingdom Child Benefit," *Journal of Human Resources*, 32, 463–480.

Nikiforidis, L., K. M. Durante, J. P. Redden, and V. Griskevicius (2018): "Do Mothers Spend More on Daughters While Fathers Spend More on Sons?" *Journal of Consumer Psychology*, 28, 149–156.

Qian, N. (2008): "Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance," *Quarterly Journal of Economics*, 123, 1251–1285.

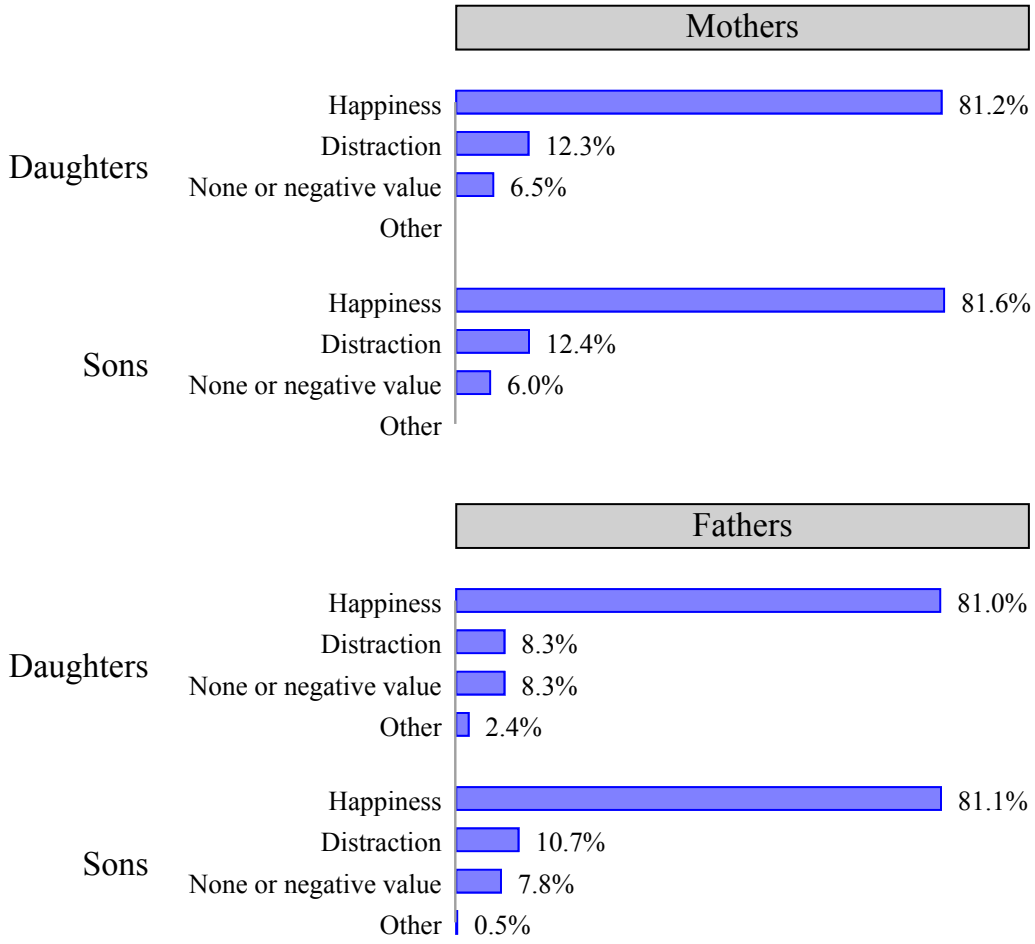
Ringdal, C. and I. H. Sjursen (2021): "HousEhold Bargaining and Spending on Children: Experimental Evidence From Tanzania," *Economica*, 88, 430–455.

Somville, V., I. Almås, and L. Vandewalle (2020): "The Effect of Gender-Targeted Transfers: Experimental Evidence From India," *NHH Dept. of Economics Discussion Paper*.

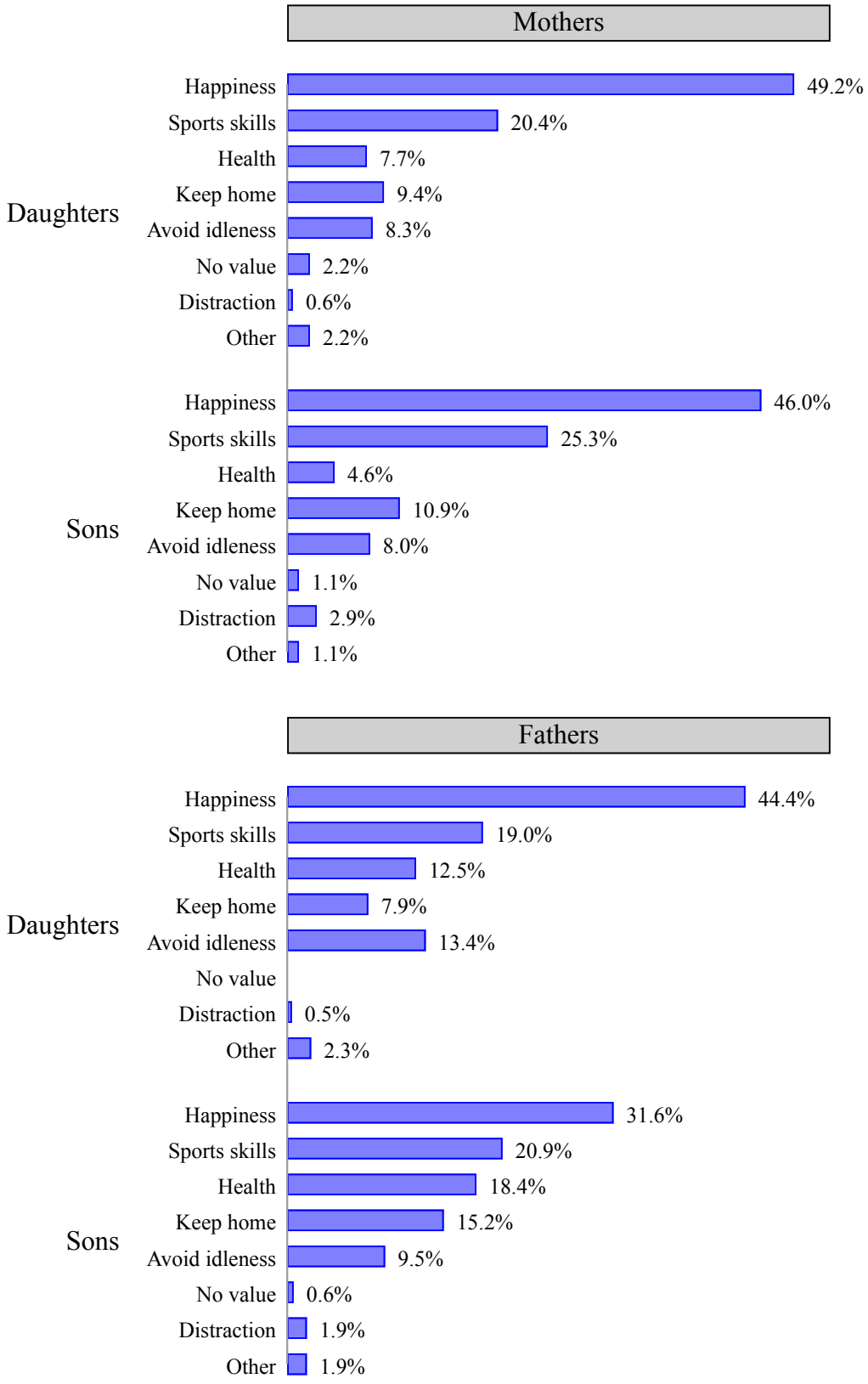
Thomas, D. (1994): "Like Father, Like Son, Like Mother, Like Daughter: Parental Education and Child Health," *Journal of Human Resources*, 29, 950–988.

A Appendix figures and tables

Appendix Figure A.1: Parents' statements of the candy's primary value, by child and parent gender



Appendix Figure A.2: Parents' statements of the rubber ball's primary value, by child and parent gender



Appendix Table A.1: Compensation for study participation does not affect WTP

	Adult good (cup) WTP	Tests WTP	Shoes WTP
	(1)	(2)	(3)
Received higher payment	-0.036	0.004	0.007
	[0.063]	[0.063]	[0.081]
Observations	1084	1084	694

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. Analysis uses first survey, in which compensation was randomized. The omitted group received 8,000 UGX for participation. The higher payment amount was 10,000 UGX; the payment level was randomized. The analysis omits deworming medicine WTP, as it was elicited only for the higher-payment group. All regressions include strata fixed effects. Standard errors are heteroskedasticity-robust.

Appendix Table A.2: Children's goods for which WTP was elicited

Good	**Survey**	**Focal child**	**Human capital?**	**Incentivized?**	**Full price (UGX)**
Practice tests	First	Older	Yes	Yes	6000
Shoes	First	Younger	Yes	No	2000/2500
Deworm	First	Older	Yes	No	4000
Workbook	Second	Older	Yes	Yes	4500
Shoes	Second	Younger	Yes	Yes	2000/2500
Deworm	Second	Younger	Yes	No	4000
Practice tests	Second	Older	Yes	No	6000
Candy	Second	Younger	No	Yes	3000
Ball	Second	Older	No	Yes	1500

Notes: The market price of shoes varied by child foot size. We used the child's age as a proxy when choosing the maximum price in the shoe WTP elicitation.

Appendix Table A.3: Household and child characteristics are balanced across mother and father samples

Variable	Mothers (1)	Fathers (2)	Standardized diff (3)
Panel A: Household characteristics			
Number of children	9.119 [3.101]	9.025 [2.967]	0.031
Number of cattle	1.006 [1.338]	1.069 [1.372]	-0.046
Number of motos	0.041 [0.199]	0.042 [0.200]	-0.005
Number of rooms	3.014 [1.197]	3.047 [1.189]	-0.028
Owns land	0.914 [0.280]	0.934 [0.248]	-0.076
Polygamous	0.247 [0.434]	0.248 [0.434]	-0.002
Panel B: Focal child characteristics			
Older focal child male	0.544 [0.498]	0.538 [0.499]	0.012
Older focal child age	11.958 [1.949]	12.047 [2.034]	-0.045
Weekly study hours	4.875 [5.305]	5.396 [9.155]	-0.070
Older focal child school performance	3.175 [0.944]	3.233 [0.939]	-0.062
Believes older focal child will support parents more than other children	0.480 [0.500]	0.495 [0.500]	-0.030
Has younger focal child	0.782 [0.413]	0.796 [0.403]	-0.034
Younger focal child male	0.553 [0.498]	0.552 [0.498]	0.002
Younger focal child age	5.733 [1.783]	5.784 [1.824]	-0.028
Younger focal child in school	0.625 [0.485]	0.675 [0.469]	-0.105
Younger focal child grade	1.488 [1.044]	1.538 [1.073]	-0.047
Number of observations	900	913	
Joint p-value	0.353		

Notes: In the regression to test for joint orthogonality, we impute missing values with the sample mean and include missing flags. We also control for survey round and strata fixed effects, to match the main specification. Standard errors clustered at the household level. The unit of observation is a household-parent.

Appendix Table A.4: Household and child characteristics are balanced across mothers and fathers within the daughter sample and within the son sample

Variable	Daughters			Sons		
	Mothers (1)	Fathers (2)	Std. diff (3)	Mothers (4)	Fathers (5)	Std. diff (6)
Panel A: Household characteristics						
Number of children	9.031 [3.058]	8.979 [2.917]	0.017	9.239 [3.050]	9.141 [2.963]	0.033
Number of cattle	0.937 [1.274]	0.995 [1.304]	-0.044	1.044 [1.352]	1.080 [1.368]	-0.027
Number of motos	0.040 [0.195]	0.035 [0.185]	0.026	0.043 [0.202]	0.043 [0.202]	0.000
Number of rooms	3.015 [1.164]	3.055 [1.160]	-0.034	2.941 [1.185]	3.021 [1.206]	-0.068
Owns land	0.912 [0.284]	0.938 [0.242]	-0.099	0.920 [0.272]	0.932 [0.252]	-0.046
Polygamous	0.242 [0.432]	0.247 [0.435]	-0.011	0.249 [0.436]	0.253 [0.438]	-0.009
Panel B: Focal child characteristics						
Older focal child male	0.820 [0.385]	0.821 [0.383]	-0.002	0.260 [0.439]	0.259 [0.438]	0.002
Older focal child age	11.827 [1.906]	11.918 [2.043]	-0.046	12.058 [1.982]	12.108 [2.003]	-0.025
Believes older focal child will support parents more than other children	0.475 [0.500]	0.497 [0.500]	-0.044	0.492 [0.500]	0.486 [0.500]	0.012
Has younger focal child	0.861 [0.346]	0.880 [0.325]	-0.059	0.895 [0.307]	0.893 [0.309]	0.006
Younger focal child male	0.335 [0.434]	0.327 [0.437]	0.017	0.771 [0.388]	0.775 [0.385]	-0.009
Younger focal child age	5.698 [1.632]	5.807 [1.713]	-0.064	5.774 [1.708]	5.756 [1.721]	0.011
Younger focal child in school	0.637 [0.410]	0.687 [0.395]	-0.122	0.625 [0.424]	0.650 [0.412]	-0.061
Younger focal child grade	1.501 [0.690]	1.545 [0.854]	-0.060	1.503 [0.733]	1.507 [0.657]	-0.005
Number of observations	805	817		799	823	
Joint p-value	0.164			0.982		

Notes: In the regression to test for joint orthogonality, we impute missing values with the sample mean and include missing flags. We also control for survey round and strata fixed effects, to match the main specification. Standard errors clustered at the household level. The unit of observation is a household-parent.

Appendix Table A.5: Household and parent characteristics are balanced across daughter and son samples

Variable	Daughters	Sons	Standardized diff
	(1)	(2)	(3)
Panel A: Household characteristics			
Number of children	9.005	9.189	-0.061
	[2.987]	[3.005]	
Number of cattle	0.966	1.062	-0.072
	[1.289]	[1.360]	
Number of motos	0.038	0.043	-0.026
	[0.190]	[0.202]	
Number of rooms	3.035	2.982	0.045
	[1.162]	[1.196]	
Owns land	0.925	0.926	-0.004
	[0.264]	[0.262]	
Polygamous	0.245	0.251	-0.014
	[0.433]	[0.437]	
Panel B: Parent characteristics			
Parent age	39.426	40.076	-0.026
	[24.740]	[24.583]	
Has some education	0.813	0.799	0.035
	[0.390]	[0.400]	
Income (10000s UGX)	122.101	121.806	0.002
	[126.532]	[134.365]	
Number of observations	1622	1622	
Joint p-value	0.240		

Notes: In the regression to test for joint orthogonality, we impute missing values with the sample mean and include missing flags. We also control for survey round and strata fixed effects, to match the main specification. Standard errors clustered at the household level. The unit of observation is a household-parent-focal child.

Appendix Table A.6: Household characteristics are balanced across daughters and sons within the mother sample and within the father sample

Variable	Mothers			Fathers		
	Daughters (1)	Sons (2)	Std. diff (3)	Daughters (4)	Sons (5)	Std. diff (6)
Panel A: Household characteristics						
Number of children	9.031 [3.058]	9.239 [3.050]	-0.069	8.979 [2.917]	9.141 [2.963]	-0.054
Number of cattle	0.937 [1.274]	1.044 [1.352]	-0.081	0.995 [1.304]	1.080 [1.368]	-0.064
Number of motos	0.040 [0.195]	0.043 [0.202]	-0.015	0.035 [0.185]	0.043 [0.202]	-0.041
Number of rooms	3.015 [1.164]	2.941 [1.185]	0.063	3.055 [1.160]	3.021 [1.206]	0.029
Owns land	0.912 [0.284]	0.920 [0.272]	-0.030	0.938 [0.242]	0.932 [0.252]	0.023
Polygamous	0.242 [0.432]	0.249 [0.436]	-0.016	0.247 [0.435]	0.253 [0.438]	-0.014
Panel B: Parent characteristics						
Age	34.900 [33.229]	35.671 [33.383]	-0.031	43.886 [9.374]	44.352 [8.526]	-0.019
Has some education	0.763 [0.426]	0.731 [0.444]	0.081	0.863 [0.344]	0.865 [0.340]	-0.005
Income (10000s UGX)	108.589 [87.413]	106.863 [85.902]	0.013	135.413 [154.657]	136.314 [167.365]	-0.007
Number of observations	805	799		817	823	
Joint p-value	0.127			0.593		

Notes: In the regression to test for joint orthogonality, we impute missing values with the sample mean and include missing flags. We omit missing flags for *Has some education* and *Parent age*, as they create a collinearity problem because the variables are only missing for 6 combined observations. We also control for survey round and strata fixed effects, to match the main specification. Standard errors clustered at the household level. The unit of observation is a household-parent-focal child.

Appendix Table A.7: Summary statistics: Mothers and fathers have different characteristics

Variable	Mothers (1)	Fathers (2)	Standardized diff (3)
Age	35.784 [31.608]	44.650 [9.258]	-0.375
Has some education	0.743 [0.437]	0.865 [0.342]	-0.307
Income (10000s UGX)	88.001 [155.807]	150.503 [211.273]	-0.320
Number of observations	899	913	

Notes: The unit of observation is a household-parent.

Appendix Table A.8: Summary statistics: Boys and girls have similar characteristics

Variable	Daughters	Sons	Standardized diff
	(1)	(2)	(3)
Panel A: All focal children			
Older focal child	0.605	0.513	0.185
	[0.489]	[0.500]	
Panel B: Older focal children			
Child age	11.873	12.083	-0.106
	[1.980]	[1.994]	
Grade	4.655	4.620	0.046
	[0.773]	[0.733]	
Weekly study hours	5.226	4.960	0.039
	[7.422]	[6.073]	
School performance	3.208	3.189	0.020
	[0.948]	[0.944]	
Will support parents more than other children	0.486	0.489	-0.006
	[0.500]	[0.500]	
Panel C: Younger focal children			
Child age	5.752	5.766	-0.008
	[1.794]	[1.813]	
In school	0.667	0.634	0.069
	[0.472]	[0.482]	
Grade	1.533	1.495	0.036
	[1.110]	[1.004]	
Number of observations	1622	1622	

Notes: The unit of observation is a household-parent-focal child.

Appendix Table A.9: Robustness of Table 1 to normalizing by mean WTP

	(1)	(2)	(3)	(4)
Daughter	-0.021	-0.056***	-0.038*	-0.047**
	[0.014]	[0.019]	[0.022]	[0.021]
Mother × Daughter		0.071***	0.044	0.072**
		[0.027]	[0.032]	[0.030]
Mother	-0.031*	-0.067***	-0.067***	-0.048**
	[0.016]	[0.021]	[0.024]	[0.024]
p-val: Mother + Mother × Daughter = 0		0.839	0.376	0.304
p-val: Daughter + Mother × Daughter = 0		0.466	0.811	0.264
Dep. var. mean father-son	1.053	1.053	1.056	1.053
Fixed effects	Stratum	Stratum	Stratum	HH
Goods included	All	All	Incentivized	All
Number of observations	6,687	6,687	4,000	6,687

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All columns control for survey round, adult WTP, adult WTP interacted with survey round, and good fixed effects. Columns 1-3 control for strata fixed effects, and column 4 controls for household fixed effects. Standard errors are clustered by household.

Appendix Table A.10: Robustness of Table 1 to controlling for *Mother* × *Younger child good*

	(1)	(2)	(3)	(4)
Daughter	-0.035	-0.102***	-0.066*	-0.099***
	[0.024]	[0.032]	[0.036]	[0.037]
Mother × Daughter		0.137***	0.073	0.152***
		[0.047]	[0.054]	[0.053]
Mother	-0.041	-0.116***	-0.087**	-0.099**
	[0.028]	[0.039]	[0.043]	[0.046]
p-val: Mother + Mother × Daughter = 0		0.580	0.750	0.224
p-val: Daughter + Mother × Daughter = 0		0.318	0.866	0.164
Dep. var. mean father-son	1.979	1.979	1.841	1.979
Fixed effects	Stratum	Stratum	Stratum	HH
Goods included	All	All	Incentivized	All
Number of observations	6,687	6,687	4,000	6,687

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All columns control for survey round, adult WTP, adult WTP interacted with survey round, and good fixed effects. Columns 1-3 control for strata fixed effects, and column 4 controls for household fixed effects. Columns 2-4 control for Mother × younger focal child good. (Note that younger focal child good is collinear with good fixed effects and so we do not need to also control for its main effect). Standard errors are clustered by household.

Appendix Table A.11: Robustness of Table 1 to omitting control for adult good WTP

	(1)	(2)	(3)	(4)
Daughter	-0.061*	-0.110**	-0.085*	-0.079**
	[0.033]	[0.044]	[0.051]	[0.039]
Mother × Daughter		0.099*	0.050	0.118**
		[0.057]	[0.069]	[0.059]
Mother	-0.052	-0.102**	-0.113**	-0.063
	[0.036]	[0.047]	[0.055]	[0.048]
p-val: Mother + Mother × Daughter = 0		0.950	0.243	0.253
p-val: Daughter + Mother × Daughter = 0		0.802	0.495	0.338
Dep. var. mean father-son	1.979	1.979	1.841	1.979
Fixed effects	Stratum	Stratum	Stratum	HH
Goods included	All	All	Incentivized	All
Number of observations	6,687	6,687	4,000	6,687

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All columns control for survey round and good fixed effects. Columns 1-3 control for strata fixed effects, and column 4 controls for household fixed effects. Standard errors are clustered by household.

Appendix Table A.12: Robustness of Table 2 to normalizing by mean WTP

	Goods included		
	Human capital	Education	Health
	(1)	(2)	(3)
Daughter	-0.058***	-0.050	-0.069**
	[0.021]	[0.033]	[0.027]
Mother × Daughter	0.076**	0.053	0.086**
	[0.030]	[0.045]	[0.038]
Mother	-0.045**	-0.038	-0.041
	[0.023]	[0.034]	[0.026]
p-val: Mother + Mother × Daughter = 0	0.164	0.625	0.119
p-val: Daughter + Mother × Daughter = 0	0.417	0.935	0.541
Dep. var. mean father-son	1.042	1.054	1.033
Number of observations	5,229	2,542	2,687

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All columns control for strata and good fixed effects, survey round, adult WTP, and adult WTP interacted with survey round. Standard errors are clustered by household.

Appendix Table A.13: Robustness of Table 2 to controlling for *Mother* × *Younger child good*

	Goods included		
	Human capital	Education	Health
	(1)	(2)	(3)
Daughter	-0.108***	-0.081	-0.138***
	[0.037]	[0.051]	[0.052]
Mother × Daughter	0.153***	0.085	0.202***
	[0.052]	[0.070]	[0.075]
Mother	-0.099**	-0.064	-0.098
	[0.043]	[0.052]	[0.095]
p-val: Mother + Mother × Daughter = 0	0.181	0.639	0.260
p-val: Daughter + Mother × Daughter = 0	0.225	0.930	0.234
Dep. var. mean father-son	2.042	1.618	2.380
Fixed effects	Stratum	Stratum	Stratum
Number of observations	5,229	2,542	2,687

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All columns control for strata and good fixed effects, survey round, adult WTP, and adult WTP interacted with survey round, and mother × younger focal child good. Standard errors are clustered by household.

Appendix Table A.14: Robustness of Table 2 to omitting control for adult good WTP

	Goods included		
	Human capital	Education	Health
	(1)	(2)	(3)
Daughter	-0.118***	-0.112*	-0.143**
	[0.045]	[0.066]	[0.060]
Mother	-0.072	-0.082	-0.063
	[0.046]	[0.063]	[0.056]
Mother × Daughter	0.119**	0.079	0.167**
	[0.058]	[0.084]	[0.081]
p-val: Mother + Mother × Daughter = 0	0.285	0.956	0.082
p-val: Daughter + Mother × Daughter = 0	0.987	0.618	0.690
Dep. var. mean father-son	2.042	1.618	2.380
Number of observations	5,229	2,542	2,687

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All columns control for strata and good fixed effects, and survey round. Standard errors are clustered by household.

Appendix Table A.15: Robustness of Figure 1(b) results to controlling for why the respondent valued the enjoyment goods

	(1)	(2)
Daughter	-0.091*	-0.085*
	[0.051]	[0.051]
Mother × Daughter	0.081	0.074
	[0.081]	[0.081]
Mother	-0.237***	-0.230***
	[0.062]	[0.062]
Main perceived benefit × good fixed effects	No	Yes
p-val: Mother + Mother × Daughter = 0	0.016	0.016
p-val: Daughter + Mother × Daughter = 0	0.866	0.859
Dep. var. mean father-son	1.754	1.754
Number of observations	1,458	1,458

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. The dependent variable is WTP for the good. All columns control for strata and good fixed effects, survey round, adult WTP, and adult WTP interacted with survey round.. The column (1) regression is the analog of Figure 1(b), and includes only the two enjoyment goods (rubber ball and candy). Column (2) additionally controls for fixed effects for the parent's main perceived benefit of the good × good fixed effects. Standard errors are clustered by household.

Appendix Table A.16: Summary statistics on mothers' and fathers' beliefs about education

Variable	Full sample (1)	Mothers (2)	Fathers (3)	Standardized diff (4)
Agree: It is useless to send girls to secondary school since they will marry	0.041 [0.198]	0.037 [0.188]	0.045 [0.207]	-0.040
Agree: Even boys who will become farmers will be better at farming if they have gone to school.	0.905 [0.293]	0.890 [0.313]	0.920 [0.271]	-0.102
Would make son finish primary if they wanted to quit	0.881 [0.323]	0.879 [0.326]	0.884 [0.321]	-0.015
[If yes to above] Would make son finish O levels if they wanted to quit	0.943 [0.231]	0.955 [0.208]	0.932 [0.252]	0.100
Would make daughter finish primary if they wanted to quit	0.855 [0.352]	0.847 [0.361]	0.864 [0.343]	-0.048
[If yes to above] Would make daughter finish O levels if they wanted to quit	0.945 [0.229]	0.942 [0.233]	0.947 [0.225]	-0.022
Number of observations	1813	900	913	

Notes: All variables are observed in both survey rounds, except the one reported in the first row, which is only available in the first round.

Appendix Table A.17: No gender-concordant norms about medical care decisions.

	Respondent decides medical care if child is sick (1)
Daughter	-0.019
	[0.055]
Mother × Daughter	0.012
	[0.078]
Mother	-0.080
	[0.051]
Dep. var. mean father-son	0.524
Number of observations	729

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. Regression additionally controls for strata fixed effects; standard errors are robust to heteroskedasticity. Dependent variable is an indicator that the respondent says they make the primary decisions about medical care when their child is sick.

B Validation of non-incentivized WTP

To assess the performance of our non-incentivized WTP elicitation, we examine WTP for practice tests, which is a good we asked the sample about in both an incentivized and non-incentivized manner. In the first round, the WTP elicitation for tests was incentivized and in the second round it was not. In addition, we asked several survey questions in both rounds that might predict demand for tests, such as perceived quality of the tests and spending on educational inputs, which we can use as potential predictors. We also use household and child characteristics as potential predictors.

Using the households surveyed in both rounds, we first use LASSO to identify the primary predictors of non-incentivized WTP and of incentivized WTP. We then use OLS to test for differences in the relationships between non-incentivized WTP and incentivized WTP and their primary predictors. To be able to conduct valid inference in the second step, we randomly split our sample in half and use one half to fit the predictive LASSO model and the second half for OLS inference on the predictive coefficients.

The evidence suggests that non-incentivized WTP performs well. We find that LASSO identifies the same predictors for non-incentivized and incentivized WTP, and that the predictive coefficients are similar. In addition, with OLS, we are unable to reject equality in the predictive relationships. We first show the LASSO results and then the OLS.

Table B.1 displays the coefficients from using LASSO in the first half of the sample to identify the two most informative predictors of incentivized WTP (column 1) and non-incentivized WTP (column 2). Notably, the table shows that LASSO selects the same primary predictors for both incentivized and non-incentivized WTP. In addition, the predictors it chooses are both intuitive and sensible: WTP for the adult good, and an indicator for the parent thinking that the tests are higher-quality than the tests offered by the child's school. The fact that LASSO picks the same predictors for both incentivized and non-incentivized WTP – and with similar predictive coefficients – is evidence that non-incentivized WTP performed well.

Next, we use the other half of the sample to conduct statistical inference on whether the relationship between WTP for tests and the two primary predictors chosen by LASSO differs for incentivized and non-incentivized WTP. We regress WTP for tests, pooled across incentivized and non-incentivized observations, on the two primary predictors and their interactions with whether a WTP observation was gathered in an incentivized or non-incentivized manner.[14] Table B.2 shows the results.

Reassuringly, neither interaction term is significant. Note that, while there is no significant difference in the predictors of WTP between incentivized and non-incentivized WTP, average incentivized WTP for the incentivized tests is lower than average non-incentivized.

[14] We estimate the following regression:

$$\begin{aligned} WTP_{ihc}^{test} =& \alpha + \beta AdultGoodWTP_{ihc} + \gamma AdultGoodWTP_{ihc} \times NonIncentivized_{ihc} + \\ & \lambda HiQuality_{ihc} + \delta HighQuality_{ihc} \times NonIncentivized_{ihc} + \nu NonIncentivized_{ihc} + \varepsilon_{ihc} \end{aligned} \tag{3}$$

where WTP_{ihc}^{test} is parent i in household h's WTP, either incentivized or non-incentivized, for tests for child c; $AdultGoodWTP_{ihc}$ is that same parent's WTP for the adult good; $HiQuality_{ihc}$ is an indicator for the parent thinking the tests were higher quality than the school's regular offering; and $NonIncentivized_{ihc}$ is an indicator that WTP was gathered in a non-incentivized way.

Appendix Table B.1: LASSO chooses the same predictors of incentivized WTP and non-incentivized WTP for tests

	Dep. var.: WTP for tests	
Variable	Incentivized	Non-incentivized
	(1)	(2)
Number of children	.	.
Number of cattle	.	.
Number of motos	.	.
Number of rooms	.	.
Owns land	.	.
Polygamous	.	.
Assets PCA	.	.
Adult Good WTP	.427	.463
Believes child is very likely to attend school	.	.
Expect child to finish primary	.	.
Would spend more on child's education than other parent	.	.
Would spend less on child's education than other parent	.	.
Tests more useful than classes	.	.
Tests more useful than workbook	.	.
Believes tests are higher quality than those school offers	.084	.017
Food fees from school	.	.
Uniform fees from school	.	.
Textbook fees from school	.	.
Spending on non-school books	.	.
Spending on extra lessons/coaching	.	.
Spending on education outside school	.	.
Supplemental expenses	.	.
Total spending on education	.	.
Total spending on education (log)	.	.
Child is male	.	.
Child age	.	.
Has younger focal child	.	.
Younger focal child male	.	.
Younger focal child age	.	.
Number of observations	364	364

Notes: Columns show coefficients from LASSO regressions that regress the WTP for tests on all of the predictors listed in the rows. Column (1) uses incentivized WTP for tests f as the dependent variable, and column (2) uses non-incentivized WTP for tests as the dependent variable. "." means that the LASSO coefficient is 0.

However, this effect is difficult to interpret: the incentivized and non-incentivized WTP were collected in different time periods (i.e., different survey waves, which occurred during different school years), and so we cannot distinguish whether this negative main effect simply reflects a time effect.

Appendix Table B.2: No significant difference in predictive coefficients for incentivized WTP and non-incentivized WTP

	(1)
Adult good WTP	0.544***
	[0.044]
Adult good WTP × Non-incentivized	0.028
	[0.068]
Believes tests high quality	0.179**
	[0.089]
Believes tests high quality × Non-incentivized	0.035
	[0.129]
Non-incentivized	-0.693***
	[0.114]
Number of observations	730

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. The dependent variable is the WTP for tests, pooled across incentivized and non-incentivized elicitations. "Believes test high quality" is an indicator for the respondent thinking the tests are higher quality than those offered by their child's school. The regression additionally controls for missing flags for adult good WTP and high-quality tests, and missing flags interacted with the non-incentivized binary. Standard errors clustered at household level.